Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

Cure Rate Model With Mismeasured Covariates Under Transformation

Yanyuan Ma[a] & Guosheng Yin[a]

[a] Yanyuan Ma is Professor, Institute of Statistics, University of Neuchâtel, Switzerland . Guosheng Yin is Assistant Professor, Department of Biostatistics, M. D. Anderson Cancer Center, University of Texas, Houston, TX 77030 . This work was supported in part by a grant from the Swiss National Science Foundation and a grant from the U.S. Department of Defense (W81XWH-05-2-0027). The authors thank the editor, the associate editor, and two anonymous referees for their insightful and constructive comments that improved the manuscript.
Published online: 01 Jan 2012.

To cite this article: Yanyuan Ma & Guosheng Yin (2008) Cure Rate Model With Mismeasured Covariates Under Transformation, Journal of the American Statistical Association, 103:482, 743-756, DOI: 10.1198/016214508000000319

To link to this article: http://dx.doi.org/10.1198/016214508000000319

Cure Rate Model With Mismeasured Covariates Under Transformation

Yanyuan MA and Guosheng YIN

Cure rate models explicitly account for the survival fraction in failure time data. When the covariates are measured with errors, naively treating mismeasured covariates as error-free would cause estimation bias and thus lead to incorrect inference. Under the proportional hazards cure model, we propose a corrected score approach as well as its generalization, and implement a transformation on the mismeasured covariates toward error additivity and/or normality. The corrected score equations can be easily solved through the backfitting procedure, and the biases in the parameter estimates are successfully eliminated. We show that the proposed estimators for the regression coefficients are consistent and asymptotically normal. We conduct simulation studies to examine the finite-sample properties of the new method and apply it to a real data set for illustration.

KEY WORDS: Cure model; Errors-in-variables problem; Proportional hazards model; Semiparametric method; Survival fraction.

1. INTRODUCTION

In oncology studies, it is often observed that a certain percentage of subjects are either cured following treatment or are unsusceptible to the event of interest and thus will never experience the failure (e.g., disease relapse). To explicitly incorporate the survival fraction for such data, cure rate models have been proposed and extensively investigated. The two-component mixture cure model (Berkson and Gage 1952) naturally separates the entire population into cured and noncured subjects,

$$S(t|\mathbf{X}) = \theta(\mathbf{X}) + \{1 - \theta(\mathbf{X})\}S^*(t|\mathbf{X}),$$

where $\theta(\mathbf{X})$ is the cure probability and $S^*(t|\mathbf{X})$ is a proper survival function for the uncured population, that is, $\lim_{t\to\infty} S^*(t|\mathbf{X}) = 0$. Its intuitive structure and ease of interpretation has made this mixture cure model the focus of much attention (see, e.g., Gray and Tsiatis 1989; Sposto, Sather, and Baker 1992; Laska and Meisner 1992; Kuk and Chen 1992; Maller and Zhou 1996; Sy and Taylor 2000; Lu and Ying 2004; Li, Tiwari, and Guha 2007). But the mixture cure model lacks certain desirable properties, as pointed out by Chen, Ibrahim, and Sinha (1999). Moreover, the numerical computation can be quite challenging due to the additive structure of the cured and uncured components.

Alternatively, the proportional hazards cure rate model developed by Yakovlev and Tsodikov (1996) and Tsodikov (1998a) integrates the survival times of the cured and noncured subjects into one single formulation of the survival function,

$$S(t|\mathbf{X}) = \exp\{-\theta(\mathbf{X})F(t)\}, \tag{1}$$

where $\theta(\mathbf{X})$ is a known link function and $F(t)$ is an unknown baseline cumulative distribution function (cdf). The corresponding cure rate is $S(\infty|\mathbf{X}) = \exp\{-\theta(\mathbf{X})\}$, and the hazard function is $\lambda(t|\mathbf{X}) = \theta(\mathbf{X})f(t)$, where $f(t) = dF(t)/dt$. When $\theta(\mathbf{X}) = \exp(\mathbf{X}^T\boldsymbol{\beta})$ and $\boldsymbol{\beta}$ contains an intercept b, model (1) becomes the usual Cox proportional hazards model (Cox 1972) subject to the restriction of a bounded baseline cumulative hazard function, given by $\Lambda_0(t) = F(t)\exp(b)$. Thus a cure rate model has a bounded cumulative hazard, leading to an improper survival function [i.e., $S(\infty|\mathbf{X}) > 0$], whereas a noncure model, such as the Cox model, has an unbounded cumulative hazard, thus resulting in a proper survival function [i.e., $S(\infty|\mathbf{X}) = 0$]. Yakovlev and Tsodikov (1996) and Chen et al. (1999) provided a sound biological derivation for model (1), and Tsodikov, Ibrahim, and Yakovlev (2003) provided a comprehensive review.

In reality, it is often the case that the covariate $\mathbf{X}$ can be measured only approximately or indirectly, leading to an errors-in-variables problem. If covariates with measurement errors are naively taken as error-free, then severe bias can be induced in the parameter estimates. Fuller (1987) and Carroll, Ruppert, Stefanski, and Crainiceanu (2006) explored various methods for correcting the bias. The observed variable, denoted by $\mathbf{W}$, is typically related to the true covariate $\mathbf{X}$ through a model $p_{\mathbf{W}|\mathbf{X}}(\mathbf{W}|\mathbf{X}, \boldsymbol{\xi})$, where $\boldsymbol{\xi}$ can be an unknown parameter. It is common to assume a normal additive error structure, that is, $\mathbf{W}$ equals $\mathbf{X}$ plus a normal random noise. When this normality assumption does not hold, one needs to either adapt the methodology to treat the nonnormal error or transform the covariates $\mathbf{X}$ and $\mathbf{W}$ into a normal error form (Nusser, Carriquiry, Dodd, and Fuller 1996; Eckert, Carroll, and Wang 1997).

The Cox model with measurement errors, has been studied extensively in, for example, the induced partial likelihood approach (Prentice 1982); joint models of survival times and longitudinal covariates measured with errors (Tsiatis, DeGruttola, and Wulfsohn 1995; Wulfsohn and Tsiatis 1997; Tsiatis and Davidian 2001), the regression calibration method (Wang, Hsu, Feng, and Prentice 1997), pseudo–partial likelihood methods (Zucker 2005) and in the presence of a validation set (Zhou and Pepe 1995; Zhou and Wang 2000; Hu and Lin 2002). Hu, Tsiatis, and Davidian (1998) and Song, Davidian, and Tsiatis (2002) studied semiparametric likelihood-based methods to relax the distributional assumption on the covariates. Various correction estimators and corrected scores have been provided by Stefanski (1989), Nakamura (1990, 1992), Kong and Gu (1999), Buzas (1998), Huang and Wang (2000), Augustin (2004), Gorfine, Hsu, and Prentice (2004), and Song and Huang (2005). Moreover, measurement error problems have been addressed in other contexts: Kulich and Lin (2000) explored these problems in the additive hazards model; Cheng

Yanyuan Ma is Professor, Institute of Statistics, University of Neuchâtel, Switzerland (E-mail: *yanyuan.ma@unine.ch*). Guosheng Yin is Assistant Professor, Department of Biostatistics, M. D. Anderson Cancer Center, University of Texas, Houston, TX 77030 (E-mail: *gsyin@mdanderson.org*). This work was supported in part by a grant from the Swiss National Science Foundation and a grant from the U.S. Department of Defense (W81XWH-05-2-0027). The authors thank the editor, the associate editor, and two anonymous referees for their insightful and constructive comments that improved the manuscript.

Journal of the American Statistical Association
June 2008, Vol. 103, No. 482, Theory and Methods
DOI 10.1198/016214508000000319

and Wang (2001), linear transformation models; Li and Ryan (2004), with heterogeneous covariate errors; and Li and Lin (2000), Greene and Cai (2004), and Hu and Lin (2004), in extensions to multivariate failure time data.

Although there is a rich body of literature dealing with measurement errors in censored survival data, all the aforementioned research work cannot be directly applied to the case with a cure fraction. To the best of our knowledge, measurement error issues in semiparametric cure rate models have not been addressed to date, and this is the first work to deal with estimation and inference in this regard. We can appreciate the difficulties involved in such models due to various unspecified components and their interactions, including the distributions of the unobservable variables, the censoring distribution, and the baseline distribution function. In fact, we have found that a general semiparametric method requires one to either assume covariate-independent censoring or directly model the censoring mechanism, neither of which is considered a satisfactory approach.

This research is motivated by a recent lung cancer study, in which the objective was to assess the association of patient survival with a certain biomarker expression in the tumor cell cytoplasm. For each patient, we had either one reading or two readings of biomarker expression by different pathologists to reduce the subjectivity of the evaluation. However, neither of the two measurements of biomarker expression could be considered precise. Our interest lies in investigating the potential of the biomarker as a new prognostic marker and therapeutic target for lung cancer. Figure 1 shows the Kaplan–Meier survival curves stratified by tumor histology (adenocarcinoma or squamous cell carcinoma). After approximately 7 years of follow-up, we can see a stable plateau at the tails of the survival curves, which indicates the existence of a possible cure fraction.

In this article we consider the proportional hazards cure rate model in (1), where $\mathbf{X}$ is measured with errors. The error structure is not necessarily normal additive, and multiple measurements may exist for $\mathbf{X}$. Without making any assumptions on the distribution of $\mathbf{X}$, we propose a corrected score approach based on the nonparametric maximum likelihood estimator (NPMLE) and a new transformation on the contaminated covariates. We show that the estimators for $\boldsymbol{\beta}$ and $F(t)$ are strongly consistent and converge to a Gaussian process at a root-n rate. Due to the complex natures of such functional measurement error problem and NPMLE, the derivations of these asymptotic properties are very involved. Furthermore, the proposed nonparametric transformation on the covariates to improve the error normality and additivity is very different from the Box–Cox or spline transformation described by Eckert et al. (1997). We show that our transformation, which can be broadly applied in general measurement error problems, is effective and easy to use. On the other hand, we note that the corrected score itself can be generalized to accommodate nonnormal error structure. Due to the proportional hazard structure, the NPMLE and partial likelihood estimator are equivalent, and thus the same estimator can be derived from the partial likelihood instead of NPMLE. A more detailed discussion of this issue is given in Section 6.

The rest of the article is organized as follows. In Section 2 we introduce notation and propose a computationally effective estimation procedure when covariates are measured with normal additive errors. In Section 3 we derive the asymptotic properties of the proposed estimators for $\boldsymbol{\beta}$ and $F(t)$, and in Section 4 we propose a general transformation to handle the nonnormal and nonadditive measurement error structure. In Section 5 we report the results of simulation studies that we conducted to evaluate the finite-sample properties of the estimators, along with our application of the proposed model to the lung cancer data set. We give some concluding remarks in Section 6 and outline the technical details of the proofs of the theorems in the Appendix.

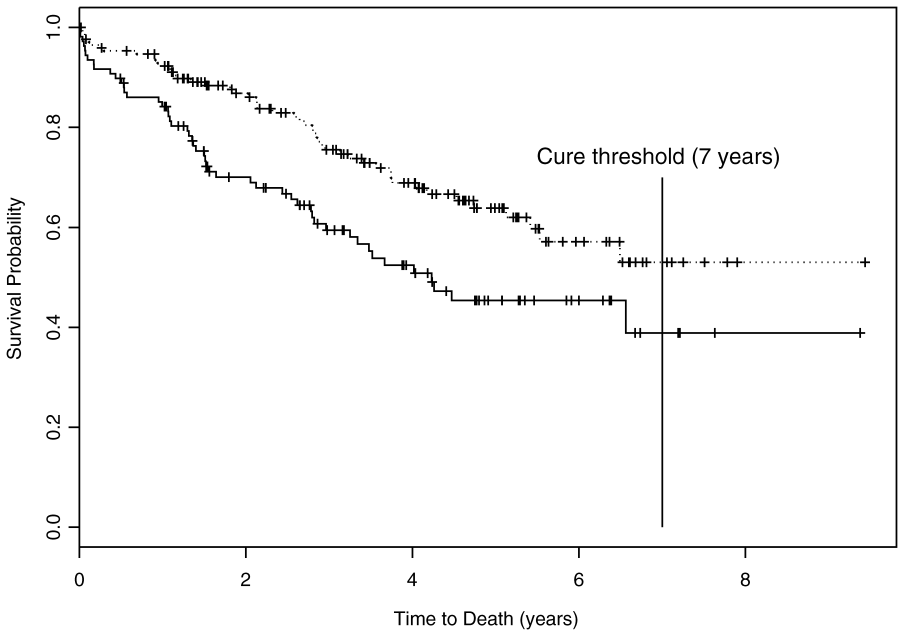

Figure 1. Estimated Kaplan–Meier survival curves for patients with lung cancer, stratified by tumor histology (—— squamous; ····· adenocarcinoma).

2. METHODOLOGY

For $i = 1, \ldots, n$, let T_i be the failure time, let C_i be the censoring time for subject i, and, correspondingly, let a q-vector $\mathbf{X}_i$ denote the covariates, the first component of which is 1. We assume that $(Y_i, \Delta_i, \mathbf{X}_i)$ are independent and identically distributed (iid), where $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$ is the censoring indicator. Furthermore, T_i is conditionally independent of C_i given covariate $\mathbf{X}_i$. The follow-up time is infinite and a proportion of subjects never experience failure or right-censoring, that is $Y_i = \infty$ with probability one for those subjects. To claim a subject cured, we need to choose a threshold and then set $Y_i = \infty$ if Y_i is larger than this threshold. In practice, a typical threshold is the largest observed event time.

For simplicity of description, we assume a classical measurement error model structure, where the error is additive and follows a mean-0 normal distribution. The treatment of the nonnormal and/or nonadditive error case is given in Section 4. We formulate the cure rate model with covariate measurement errors as

$$S(t|\mathbf{X}) = \exp\{-F(t)e^{\mathbf{X}^T\boldsymbol{\beta}}\}, \qquad \mathbf{W} = \mathbf{X} + \mathbf{U},$$

where the error $\mathbf{U} \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$. The observations are of the form $\{(Y_i, \Delta_i, \mathbf{W}_{i1}, \ldots, \mathbf{W}_{ir_i}), i = 1, \ldots, n\}$; that is, for each unobservable $\mathbf{X}_i$, we have r_i replicated observations of $\mathbf{W}_{ij}$'s $(j = 1, \ldots, r_i)$. The number of replicates is allowed to vary across different subjects, and it also may occur that $r_i = 1$. The case where some covariates are error-free is accommodated in our model by setting the relevant terms in $\mathbf{V}$ to be 0. We further make the typical surrogacy assumption that $\mathbf{W}$ and Y are independent conditional on $\mathbf{X}$. Thus, given the unobserved true covariate $\mathbf{X}$, the observed covariate $\mathbf{W}$ does not contain any additional information.

Write the survival function and the probability density function for the event of interest as $S_e(Y|\mathbf{X}) = \exp\{-F(Y)e^{\mathbf{X}^T\boldsymbol{\beta}}\}$ and $f_e(Y|\mathbf{X}) = \exp\{-F(Y)e^{\mathbf{X}^T\boldsymbol{\beta}}\}f(Y)e^{\mathbf{X}^T\boldsymbol{\beta}}$. Similarly, for the censoring times, let $S_c(Y|\mathbf{X}) = \Pr(C \ge Y|\mathbf{X})$ and $f_c(Y|\mathbf{X}) = -\partial S_c(Y|\mathbf{X})/\partial Y$. Under the cure rate model, we know that $S_e(\infty|\mathbf{X}) = \exp(-e^{\mathbf{X}^T\boldsymbol{\beta}}) > 0$ and $S_c(\infty|\mathbf{X}) > 0$.

If $\mathbf{X}$ is observed, then the likelihood of a single observation (Y, Δ) given $\mathbf{X}$ can be written as

$$\begin{aligned} &f(Y, \Delta|\mathbf{X}) \\ &\quad = \left[\{f_e(Y|\mathbf{X})S_c(Y|\mathbf{X})\}^{\Delta}\{f_c(Y|\mathbf{X})S_e(Y|\mathbf{X})\}^{1-\Delta}\right]^{I(Y<\infty)} \\ &\qquad \times \{S_e(\infty|\mathbf{X})S_c(\infty|\mathbf{X})\}^{I(Y=\infty)}. \end{aligned}$$

Similar to the work of Zeng, Yin, and Ibrahim (2006), we construct a sieve of the distribution function F, and thus the log-likelihood is given by

$$\begin{aligned} \log f(Y, \Delta|\mathbf{X}) = {} & \Delta I(Y<\infty)\{-F(Y)e^{\mathbf{X}^T\boldsymbol{\beta}} + \log F\{Y\} \\ & + \mathbf{X}^T\boldsymbol{\beta} + \log S_c(Y|\mathbf{X})\} \\ & + (1-\Delta)I(Y<\infty) \\ & \times \{\log f_c(Y|\mathbf{X}) - F(Y)e^{\mathbf{X}^T\boldsymbol{\beta}}\} \\ & + I(Y=\infty)\{\log S_c(\infty|\mathbf{X}) - e^{\mathbf{X}^T\boldsymbol{\beta}}\}, \end{aligned}$$

where $F\{Y\}$ denotes the jump size of $F(\cdot)$ at Y and $F(\cdot)$ is a right-continuous function with jumps at event times only. For ease of exposition, we write $p_i \equiv F\{Y_i\}$, denote the ordered distinct failure times as $(Y_{(1)}, \ldots, Y_{(m)})$, and denote the corresponding jump sizes as $(p_{(1)}, \ldots, p_{(m)})$, where m is the number of distinct failure times. Under the constraint $\sum_{i=1}^m p_{(i)} = 1$, we introduce a Lagrange multiplier, λ, and maximize

$$\sum_{i=1}^n \log f(Y_i, \Delta_i|\mathbf{X}_i) - n\lambda\left(\sum_{i=1}^m p_{(i)} - 1\right)$$

with respect to $(\boldsymbol{\beta}, \lambda, p_{(1)}, \ldots, p_{(m)})$. Collecting only the terms containing the unknown parameters $(\boldsymbol{\beta}, \lambda, F)$, this is equivalent to maximizing

$$\begin{aligned} &\sum_{i=1}^n \left[-F(Y_i)e^{\mathbf{X}_i^T\boldsymbol{\beta}} + \Delta_i I(Y_i<\infty)(\log F\{Y_i\} + \mathbf{X}_i^T\boldsymbol{\beta})\right] \\ &\qquad - n\lambda\left(\sum_{i=1}^m p_{(i)} - 1\right), \end{aligned} \tag{2}$$

where $F(Y_i) = \sum_{Y_j \le Y_i, \Delta_j = 1} F\{Y_j\}$ and $F(\infty) = 1$.

As opposed to using a profile likelihood approach (see Zeng et al. 2006), we take a backfitting procedure to maximize the log-likelihood. To be more specific, we solve for the $p_{(i)}$'s and λ by fixing $\boldsymbol{\beta}$, and solve for $\boldsymbol{\beta}$ by fixing the $p_{(i)}$'s and λ. The derivatives of (2) with respect to the $p_{(i)}$'s and λ are

$$\frac{1}{p_{(i)}} = \sum_{j=1}^n I\left(Y_{(i)} \le Y_j < \infty\right)e^{\mathbf{X}_j^T\boldsymbol{\beta}} + n\lambda, \qquad i = 1, \ldots, m, \tag{3}$$

and

$$\sum_{i=1}^m p_{(i)} = 1. \tag{4}$$

Therefore, we can iterate between (3), (4), and

$$\sum_{i=1}^n \left\{\Delta_i I(Y_i<\infty) - F(Y_i)e^{\mathbf{X}_i^T\boldsymbol{\beta}}\right\}\mathbf{X}_i = \mathbf{0} \tag{5}$$

to obtain the estimators.

When $\mathbf{X}$ is not observable but the $\mathbf{W}$'s are observed instead, we modify the estimating equations so that they are functions of the observed data and yield consistent estimators. We keep (4) unchanged. Following the corrected score approach, we modify the m equations in (3) by replacing $e^{\mathbf{X}_i^T\boldsymbol{\beta}}$ with $e^{\mathbf{W}_i^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}$,

$$\begin{aligned} \frac{1}{p_{(i)}} = \sum_{j=1}^n \frac{1}{r_j}\sum_{k=1}^{r_j} I\left(Y_{(i)} \le Y_j < \infty\right)e^{\mathbf{W}_{jk}^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2} + n\lambda, \\ i = 1, \ldots, m. \end{aligned} \tag{6}$$

An alternative way to handle multiple measurements is to take an average of the $\mathbf{W}_{ik}$'s for each i a priori to form a single "better" observation, that is, using

$$\frac{1}{p_{(i)}} = \sum_{j=1}^n I\left(Y_{(i)} \le Y_j < \infty\right)e^{\bar{\mathbf{W}}_j^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{V}_j\boldsymbol{\beta}/2} + n\lambda$$

to replace (3), where $\bar{\mathbf{W}}_i = {r_i}^{-1}\sum_{k=1}^{r_i}\mathbf{W}_{ik}$ and $\mathbf{V}_i = r_i^{-1}\mathbf{V}$. In practice, we have found that the two treatments of the $\mathbf{W}_{ik}$'s

yield very similar results. To modify (5), we replace $\mathbf{X}_i$ with $\mathbf{W}_i$ and $e^{\mathbf{X}_i^T\boldsymbol{\beta}}\mathbf{X}_i$ with $e^{\mathbf{W}_i^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}(\mathbf{W}_i-\mathbf{V}\boldsymbol{\beta})$ to obtain

$$\sum_{i=1}^{n}\frac{1}{r_i}\sum_{k=1}^{r_i}\{\Delta_i I(Y_i<\infty)\mathbf{W}_{ik} - F(Y_i)e^{\mathbf{W}_{ik}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}(\mathbf{W}_{ik}-\mathbf{V}\boldsymbol{\beta})\}=\mathbf{0}. \quad (7)$$

The final estimators under the corrected scores can be obtained by solving (4), (6), and (7) simultaneously.

3. ASYMPTOTIC PROPERTIES

For ease of exposition, we focus on the situation in which we have one surrogate $\mathbf{W}$ for the true unobserved $\mathbf{X}$. We first introduce some notation and define $l^\infty(\mathcal{H})$ as the space of all bounded linear functionals on $\mathcal{H}$, where

$$\mathcal{H}=\{(\mathbf{h}_1,h_2):\mathbf{h}_1\in\mathcal{R}^q,\|\mathbf{h}_1\|<1, h_2 \text{ is a function in } [0,\infty) \text{ with } \|h_2\|_V\le 1\}$$

and $\|h_2\|_V$ is the total variation of h_2. Let $l(\boldsymbol{\beta},F)$ be the log-likelihood conditional on $\mathbf{X}$,

$$l(\boldsymbol{\beta},F)=-F(Y)e^{\mathbf{X}^T\boldsymbol{\beta}}+\Delta I(Y<\infty)\{\log f(Y)+\mathbf{X}^T\boldsymbol{\beta}\},$$

where we omit the part of $l(\boldsymbol{\beta},F)$ that does not involve $\boldsymbol{\beta}$ or F. Denote the derivative of $l(\boldsymbol{\beta},F)$ with respect to $\boldsymbol{\beta}$ as $\mathbf{l}_{\boldsymbol{\beta}}(\boldsymbol{\beta},F)$ and write $l_F(\boldsymbol{\beta},F)[\int Q_F(h_2)\,dF]$ as the derivative of $l(\boldsymbol{\beta},F)$ along the path $(\boldsymbol{\beta},F_\epsilon=F+\epsilon\int Q_F(h_2)\,dF),\epsilon\in(-\epsilon_0,\epsilon_0)$ for a small constant $\epsilon_0>0$, where $Q_F(h_2)=h_2(t)-\int_0^\infty h_2(t)\,dF(t)$. This operation ensures that $Q_F(h_2)$ integrates to 0, and thus the perturbed F remains a valid cdf. This restriction plays the same role as the Lagrange multiplier. The corrected scores can be constructed by replacing the terms involving $\mathbf{X}$ in $\mathbf{l}_{\boldsymbol{\beta}}(\boldsymbol{\beta},F)$ and $l_F(\boldsymbol{\beta},F)[\int Q_F(h_2)\,dF]$ with those involving $\mathbf{W}$. We denote the expressions after such replacement as $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta},F)$ and $S_F(\boldsymbol{\beta},F)[\int Q_F(h_2)\,dF]$. Note that $\int Q_F(h_2)\,dF$ does not involve $\mathbf{X}$ or $\mathbf{W}$. Straightforward calculation yields that

$$\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta},F)=-F(Y)e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})+\Delta I(Y<\infty)\mathbf{W}$$

and

$$S_F(\boldsymbol{\beta},F)\left[\int Q_F(h_2)\,dF\right] = -\int_0^Y Q_F(h_2)\,dF(t)\,e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2} + \Delta I(Y<\infty)Q_F\{h_2(Y)\}.$$

Let $\mathbb{P}_n$ and $\mathbb{P}$ denote the empirical measure of n iid observations and the expectation; that is, for any measurable function $g(Y,\Delta,\mathbf{X})$ in $L_2(P)$,

$$\mathbb{P}_n[g(Y,\Delta,\mathbf{X})]=\frac{1}{n}\sum_{i=1}^{n}g(Y_i,\Delta_i,\mathbf{X}_i) \qquad \text{and}$$

$$\mathbb{P}[g(Y,\Delta,\mathbf{X})]=E[g(Y,\Delta,\mathbf{X})].$$

We assume that our model is identifiable and that the following regularity conditions are satisfied:

(C1) The covariate $\mathbf{W}$ is bounded with probability 1.

(C2) Conditional on $\mathbf{X}$, the censoring time C is independent of T, and $P(C=\infty|\mathbf{X})>0$.

(C3) The true parameter $\boldsymbol{\beta}_0$ belongs to the interior of a known compact set $\mathcal{B}_0$, and the true cdf F_0 is differentiable with its first derivative $f_0(t)>0$ for all $t\in\mathcal{R}^+$.

These are rather mild conditions that are routinely made in cure rate models. We now present the asymptotic properties of the estimators, including strong consistency, asymptotic normality, and the variance estimation formula.

Theorem 1. Under the regularity conditions, assume the limiting estimating equation

$$\mathbb{P}\left\{\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta},F)^T\mathbf{h}_1+\mathbf{S}_F(\boldsymbol{\beta},F)\left[\int Q_F(h_2)\,dF\right]\right\}=0$$

has a unique zero. With probability 1, the estimators $\hat{\boldsymbol{\beta}}_n$ and $\hat{F}_n(t)$ of (4), (6), and (7) satisfy

$$|\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0|\to\mathbf{0} \qquad \text{and} \qquad \sup_{t\in\mathcal{R}^+}|\hat{F}_n(t)-F_0(t)|\to 0.$$

Theorem 2. Under the regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0,\hat{F}_n-F_0)$ converges weakly to a mean-0 Gaussian process in $l^\infty(\mathcal{H})$.

The proofs of these two theorems depend heavily on the empirical process theory (van der Vaart and Wellner 2000), which are outlined in the Appendix.

Theorem 3. Under the regularity conditions, the estimator $\hat{\boldsymbol{\beta}}_n$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)\to\mathrm{N}(\mathbf{0},\mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^T)$$

in distribution as $n\to\infty$, where

$$\mathbf{A}=E\left(e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}\left[F_0(Y)\{\mathbf{V}-(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})^T\} - (\mathbf{W}-\mathbf{V}\boldsymbol{\beta})\int_0^Y\mathbf{b}_4(y)^T\,dF_0(y)\right]\right)^T,$$

$$\mathbf{B}=\left\{\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0,F_0)+S_F(\boldsymbol{\beta}_0,F_0)\left[\int_0^Y\mathbf{b}_4(y)\,dF_0(y)\right]\right\}^{\otimes 2},$$

and $\mathbf{b}_4$ is given in (A.5) in the Appendix.

The derivation of the variance sandwich formula with covariate measurement errors is very different from that without measurement errors, as shown in the Appendix.

4. NONNORMAL AND NONADDITIVE ERROR

Measurement error models often require transforming covariates toward the error normality and additivity. Considering one component of $\mathbf{W}$ and $\mathbf{X}$, we need to find a suitable transformation function ϕ so that $\phi(W)=\phi(X)+U$, where U follows a mean-0 normal distribution. When the transformation ϕ belongs to a parametric family indexed by $\boldsymbol{\gamma}$, for the case with duplicates ($r_i=2$), we have that

$$\phi(W_1,\boldsymbol{\gamma})=\phi(X,\boldsymbol{\gamma})+U_1, \qquad \phi(W_2,\boldsymbol{\gamma})=\phi(X,\boldsymbol{\gamma})+U_2,$$

where U_1 and U_2 are independent mean-0 normal variables. We can estimate the parameter $\boldsymbol{\gamma}$ through the maximum likelihood approach based on $\{\phi(W_{i1},\boldsymbol{\gamma})-\phi(W_{i2},\boldsymbol{\gamma})\}/\sqrt{2}$, for

$i = 1, \ldots, n$. But in many practical situations, a standard transformation family, such as the Box–Cox or power transformation, may not be sufficient to achieve the desired normal and additive error structure. Eckert et al. (1997) proposed a class of transformations based on piecewise cubic spline functions. This family of transformation is nonparametric and versatile and often performs superior to the Box–Cox transformation. However, its implementation can be difficult, and the transformation contains several ad hoc procedures.

To enhance the flexibility of our model, we propose a transformation that is easy to use and completely data-driven. We first sort all of the W_{ij}'s in an increasing order, denoted as $(W_{(1)}, W_{(2)}, \ldots, W_{(2n)})$. Let $b_i = \phi(W_{i1}) - \phi(W_{i2})$, $b_{(i)}$'s be the order statistics, and let $(q_1, q_2, \ldots, q_n)$ correspond to the $(.5/n, 1.5/n, \ldots, 1 - .5/n)$ quantiles of the standard normal distribution. Then we search for the set of $\phi(W_{ij})$'s that minimizes $\sum_{i=1}^{n}[\{b_{(i)} - \mu_b\}/\sigma_b - q_i]^2$, where μ_b and σ_b^2 are the sample mean and sample variance of the b_i's. The minimization is performed under a monotonic constraint that the $\phi(W_{ij})$'s follow exactly the same order as the W_{ij}'s, that is, $\phi(W_{(1)}) \le \phi(W_{(2)}) \le \cdots \le \phi(W_{(2n)})$. The basic intuition behind this operation is that we want to find a transformation ϕ so that the resulting n sample quantiles are the closest to the expected quantiles in terms of the mean squared error (MSE). Other than monotonicity, we impose no constraints on ϕ; thus this transformation is more flexible than those proposed in the literature. For convenience, we use the MSE as the evaluation criterion for the transformation; one certainly could opt for other criteria, such as the mean absolute deviation or a weighted average, which may emphasize the central part more than the tail part of the data or may focus only on the maximum distance.

The proposed transformation is rank-preserving but cannot make a distinction between ϕ and $a_0 + a_1\phi$ for any constants a_0 and a_1. Thus, to ensure identifiability, we set the first two values, $\phi(W_{(1)})$ and $\phi(W_{(2)})$, to two constants, say $\phi(W_{(1)}) = 1$ and $\phi(W_{(2)}) = 2$. Our original problem involves an order-constrained minimization, which often requires rather specialized optimization routines. If we reparameterize and take $\phi(W_{(i)}) = \sum_{j=1}^{i} e^{\tau_j}$, then we can minimize $\sum_{i=1}^{n}[\{b_{(i)} - \mu_b\}/\sigma_b - q_i]^2$ without constraints to obtain the τ_j's and hence the $\phi(W_{ij})$'s. Note that a single value change in any of the τ_j's would cause changes in μ_b and σ_b; thus the optimization cannot be simplified by investigating each individual term separately. In addition, the fixed order of $\phi(W_{ij})$'s does not imply a fixed order of b_i's; thus the objective function may not be differentiable at the τ_j values at which a change in the order of b_i's occurs. Because of these considerations, we use a large-scale Nelder–Mead simplex method as the optimization procedure, in combination with multiple sets of dispersed starting values for τ_j's, to avoid convergence to local minima. It is worth pointing out that although the optimal solution gives the best transformation toward normality, in reality, we would be content as long as the resulting $\phi(W_{i1}) - \phi(W_{i2})$ was sufficiently close to normality. Various procedures can be used to examine the performance of the transformation. We formulate the Pearson-type statistic in the form of $\sum_{k=1}^{K}(E_k - O_k)^2/E_k$, where K is the number of partitions of the data space, and E_k and O_k are the expected and observed bin counts. Under the null model in which $(b_i - \mu_b)/\sigma_b$ follows the standard normal distribution, the Pearson statistic asymptotically follows a chi-squared distribution with degrees of freedom $K - 1$ (see, e.g., Rao 1973). The proposed nonparametric transformation is straightforward and is quite effective, as we show in our numerical studies. Once we obtain the $\phi(W_{ij})$'s, the variance V can be easily estimated using the sample variance of $\{\phi(W_{i1}) - \phi(W_{i2})\}/\sqrt{2}$, $i = 1, \ldots, n$.

In general, there is no guarantee that a normal additive error can always be achieved. In cases where the normal error cannot be obtained, the estimating equations (3) and (5) should be corrected by replacing $\mathbf{X}_j$, $e^{\mathbf{X}_j^T\boldsymbol{\beta}}$ with $\mathbf{W}_j - E(\mathbf{U}_j)$, $e^{\mathbf{W}_j^T\boldsymbol{\beta}}/E(e^{\mathbf{U}_j^T\boldsymbol{\beta}})$ and $\mathbf{X}_j e^{\mathbf{X}_j^T\boldsymbol{\beta}}$ with $\mathbf{W}_j e^{\mathbf{W}_j^T\boldsymbol{\beta}}/E(e^{\mathbf{U}_j^T\boldsymbol{\beta}}) - e^{\mathbf{W}_j^T\boldsymbol{\beta}}E(\mathbf{U}_j e^{\mathbf{U}_j^T\boldsymbol{\beta}})/E(e^{\mathbf{U}_j^T\boldsymbol{\beta}})^2$.

5. NUMERICAL STUDIES

5.1 Simulation

We conducted three sets of simulation studies to examine the small-sample performance of the proposed methods. First, we studied a cure rate model function,

$$S(t|X_1, X_2) = \exp\{-\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)F(t)\}, \quad (8)$$

where X_1 is a uniformly distributed random variable on $[0, 1]$ and subject to measurement errors and X_2 is a Bernoulli random variable that takes a value of 0 or 1 with equal probability. We took the true parameters $\beta_0 = .5$, $\beta_1 = 1$, $\beta_2 = -.5$, and $F(t) = 1 - \exp(-t)$. The measurement error model was formulated as $W = X_1 + U$, where U was a normal random variable with mean 0 and standard deviation σ. We considered $\sigma = .1$ and .2 to examine the impact of the measurement error on the estimators. When the censoring time was generated from an exponential distribution with mean 1, designated as exp(1), the resulting data set had an approximate censoring rate of 17%, and a cure rate of 8%; and when the censoring time was generated from exp(.1), it yielded a censoring rate of 33%. We took sample sizes of $n = 200$ and 300, and performed 1,000 simulations under each configuration. For each data replicate, we implemented the backfitting procedure to estimate β_0, β_1, and β_2 and the corresponding variances. The corrected estimating equations were solved using the Newton–Raphson algorithm, which converged very fast and was quite robust to the initial values. For comparison, we also carried out a naive estimation procedure, in which the measurement error was ignored and W was treated as X_1. The simulation results are presented in Tables 1 and 2, corresponding to censoring rates of 17% and 33%. As we can see, even with a small measurement error scale, $\sigma = .1$, the naive estimator of β_0 was biased upward and that of β_1 was biased downward, and these biased increased severely as the measurement error increased to $\sigma = .2$. The corresponding coverage probabilities of 95% confidence intervals were under the nominal level, especially for the cases with $\sigma = .2$. Interestingly, because covariate X_2 was measured precisely, the estimator of β_2 under the naive method performed well; the bias was negligible, and the coverage probability was close to 95%. In contrast, the proposed estimator successfully corrected the bias under all of the scenarios. Moreover, the estimated variances based on the asymptotic normal approximation formula

Table 1. Simulation results under model (8) with 17% censoring

			Proposed method			Naive method		
σ	n	Estimate	β_0	β_1	β_2	β_0	β_1	β_2
.1	200	Bias	−.011	.031	.002	.041	−.089	.006
		Empirical variance	.039	.095	.031	.035	.072	.030
		Estimated variance	.037	.098	.028	.031	.073	.027
		95% cv	.941	.956	.936	.923	.927	.929
	300	Bias	−.005	.021	−.001	.046	−.096	.003
		Empirical variance	.028	.066	.019	.025	.050	.018
		Estimated variance	.025	.064	.018	.020	.048	.018
		95% cv	.932	.939	.956	.901	.926	.950
.2	200	Bias	−.007	.028	−.003	.152	−.336	.007
		Empirical variance	.054	.164	.032	.033	.058	.030
		Estimated variance	.052	.156	.040	.026	.055	.027
		95% cv	.941	.946	.944	.805	.687	.936
	300	Bias	−.008	.026	−.004	.148	−.331	.005
		Empirical variance	.034	.103	.020	.022	.038	.018
		Estimated variance	.033	.100	.019	.017	.037	.018
		95% cv	.943	.952	.951	.764	.576	.950

NOTE: 95% CV represents the coverage probability of 95% confidence intervals.

were quite close to the empirical variances, and our method produced satisfactory coverage probabilities at the 95% nominal level. When σ increased, the estimated variances for the β's increased as more variability was incorporated into the model and the estimation procedure. However, the opposite was true for the naive estimators. Because W was treated as the true covariate X_1, more variation in the covariate would produce better estimators; the variances using the naive method were in fact smaller for $\sigma = .2$ compared with those with $\sigma = .1$. At a higher censoring rate, as shown in Table 2, similar conclusions can be drawn. The estimated variances increased as the censoring percentage increased and decreased as the sample size grew large.

Our second simulation was designed to study a scenario with replicates for mismeasured covariates. We considered a cure rate model,

$$S(t|X_1, X_2, X_3, X_4) = \exp\{-\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4) F(t)\}, \quad (9)$$

where both X_1 and X_2 are Bernoulli random variables that take a value of 1 with probabilities of .5 and .6, and X_3 and X_4 are generated from uniform distributions on $[-.5, .5]$ and $[0, 1]$. Here X_4 was unobservable; instead, we observed two replicates (W_1, W_2), where each replicate was X_4 plus a normal error with mean 0 and standard deviation $\sigma = .2$. The true parameters were $\beta_0 = .5$, $\beta_1 = .5$, $\beta_2 = -.5$, $\beta_3 = 1$, $\beta_4 = -1$, and $F(t) = 1 - \exp(-t)$. The censoring times were generated independently from exp(1), yielding an approximate censoring

Table 2. Simulation results under model (8) with 33% censoring

			Proposed method			Naive method		
σ	n	Estimate	β_0	β_1	β_2	β_0	β_1	β_2
.1	200	Bias	−.006	.016	−.003	.046	−.104	.000
		Empirical variance	.054	.135	.038	.047	.102	.037
		Estimated variance	.050	.126	.036	.041	.093	.034
		95% cv	.949	.946	.937	.928	.924	.935
	300	Bias	−.005	.006	.003	.046	−.110	.006
		Empirical variance	.035	.085	.025	.031	.065	.024
		Estimated variance	.033	.082	.024	.028	.062	.023
		95% cv	.939	.950	.945	.917	.921	.944
.2	200	Bias	−.025	.078	−.008	.141	−.312	.007
		Empirical variance	.063	.207	.042	.038	.071	.038
		Estimated variance	.070	.205	.038	.036	.070	.034
		95% cv	.958	.951	.938	.863	.776	.934
	300	Bias	−.001	.032	−.009	.155	−.331	.003
		Empirical variance	.048	.149	.026	.030	.055	.024
		Estimated variance	.044	.128	.025	.023	.047	.023
		95% cv	.943	.932	.946	.790	.646	.950

NOTE: 95% CV represents the coverage probability of 95% confidence intervals.

Table 3. Simulation results under model (9) with 25% censoring

		Proposed method					Naive method				
n	Estimate	β_0	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
Average $\bar{W}$											
200	Bias	.021	.009	−.013	.028	−.037	−.091	.005	−.008	.020	.184
	Empirical variance	.078	.044	.042	.134	.182	.060	.042	.040	.130	.108
	Estimated variance	.069	.042	.042	.128	.173	.053	.041	.040	.123	.099
	95% cv	.931	.952	.953	.946	.948	.920	.950	.953	.946	.896
300	Bias	.012	.002	−.006	.013	−.023	−.096	−.002	−.001	.006	.189
	Empirical variance	.051	.030	.028	.082	.115	.039	.030	.027	.081	.069
	Estimated variance	.045	.028	.027	.084	.110	.035	.027	.027	.081	.065
	95% cv	.926	.948	.950	.960	.953	.912	.947	.952	.958	.869
Replicates (W_1, W_2)											
200	Bias	.023	.009	−.012	.028	−.040	−.162	.002	−.005	.014	.325
	Empirical variance	.080	.044	.042	.135	.187	.051	.042	.039	.129	.073
	Estimated variance	.069	.042	.042	.128	.182	.046	.041	.040	.122	.068
	95% cv	.929	.950	.955	.944	.956	.877	.954	.954	.945	.745
300	Bias	.012	.002	−.006	.013	−.024	−.166	−.005	.002	.001	.327
	Empirical variance	.052	.030	.028	.082	.117	.034	.029	.027	.080	.047
	Estimated variance	.045	.028	.027	.084	.114	.030	.027	.026	.081	.045
	95% cv	.926	.946	.951	.959	.952	.835	.945	.953	.956	.654

NOTE: 95% CV represents the coverage probability of 95% confidence intervals.

rate of 25%. We implemented two different treatments of the replicates: averaging W_1 and W_2 to obtain a single "better" measurement, $\bar{W} = (W_1 + W_2)/2$, and incorporating each individual measurement (W_1, W_2) in the estimation as in (6) and (7). The simulation results, given in Table 3, show that the estimates using $\bar{W}$ or (W_1, W_2) were comparable. In particular, the bias could be corrected satisfactorily compared with the naive method, the asymptotic variance provided a good approximation of the empirical variance, and the 95% coverage probability closely matched the nominal level. As the sample size increased, the bias and variance decreased. But the naive method had obvious biases in the estimates of β_0 and β_4, whereas the parameter estimates for the precisely measured covariates X_1, X_2, and X_3 were satisfactory. Furthermore, it is interesting that using the duplicates (W_1, W_2) led to much worse biases and coverage probabilities than those resulting from using the average $\bar{W}$ based on the naive method. The averaged covariate values could offset the effect of measurement errors to a certain extent, because the random noise would diminish by averaging over multiple replicates.

In the third simulation, we conducted a sensitivity analysis to demonstrate the effectiveness of the proposed transformation and the robustness of our model to the misspecified normal additive error. We examined model (8) with measurement error structure $W = \exp(X_1 + U)$, where we generated U from a normal, a Student t with degrees of freedom 10 and 5, and a uniform distribution. We set the mean of U to 0 and the standard deviation of U to .2 for all of the scenarios. We implemented the proposed transformation on the covariate, even though for all cases but the first, whether or not a normal additive error structure could be obtained is not clear. Table 4 shows that the proposed method performed well by imposing our transformation when W is linked to X through a nonnormal and nonadditive error. We could capture the true transformation that recovered the normal additive error structure. When the normality assumption was violated, our estimation procedure appeared to be quite robust and still produced estimates with very small biases. As the degree of freedom of the t distribution decreased to 5, the performance deteriorated slightly. In the simulations not reported here, we also explored other transformations, such as $W = (X_1 + U)^3$, and found similar results.

In the foregoing simulations, the censoring distribution has an infinite support. Because in reality the censoring time is always finite, we also conducted simulations in which the censoring distribution was finitely supported. Here a subject is considered cured if the subject is censored and the corresponding censoring time is larger than the largest observed event time. The estimator remains consistent, and the variance estimation and the 95% coverage probability are satisfactory as well.

5.2 Lung Cancer Data

As an illustration, we applied the cure rate model with measurement errors to the lung cancer data set. The study group comprised 280 patients. The covariates of interest included either one or two readings of biomarker expression, tumor histology (61% adenocarcinoma = 1; 39% squamous cell carcinoma = 0), and patient age (range, 34 to 90 years; mean, 66 years) and sex (52% female = 1; 48% male = 0). The covariate age was standardized to have mean 0 and variance 1. The underlying true expression of the biomarker could not be measured precisely. For half of the patients, only one reading of biomarker expression was available, whereas for the other half, two different readings were recorded, with no preference given to either reading.

For the 140 patients with 2 readings of biomarker expression, we took the difference of the logarithm of the 2 readings and found that the original observations of biomarker expression did not satisfy the normal additive error structure. After carrying out our transformation on the readings of biomarker

Table 4. Transformation and sensitivity analysis of the error structure under model (8)

			Proposed method			Naive method		
n	Error	Estimate	β_0	β_1	β_2	β_0	β_1	β_2
200	Normal	Bias	−.017	.047	−.003	.147	−.334	.008
		Empirical variance	.055	.163	.032	.034	.057	.029
		Estimated variance	.052	.159	.030	.026	.056	.028
		95% cv	.952	.951	.938	.808	.695	.943
	t_{10}	Bias	−.015	.044	−.006	.147	−.334	.009
		Empirical variance	.057	.173	.033	.035	.061	.030
		Estimated variance	.051	.159	.030	.026	.056	.028
		95% cv	.939	.942	.946	.812	.690	.946
	t_5	Bias	−.084	.203	−.011	.120	−.273	.007
		Empirical variance	.066	.220	.033	.036	.067	.030
		Estimated variance	.062	.203	.031	.027	.061	.028
		95% cv	.948	.939	.946	.847	.780	.947
	Uniform	Bias	−.017	.047	−.003	.145	−.329	.007
		Empirical variance	.054	.150	.031	.035	.055	.029
		Estimated variance	.052	.158	.030	.026	.057	.028
		95% cv	.954	.957	.952	.803	.718	.944
300	Normal	Bias	−.006	.017	−.001	.150	−.341	.009
		Empirical variance	.035	.106	.020	.022	.041	.019
		Estimated variance	.033	.100	.019	.017	.037	.018
		95% cv	.936	.944	.951	.762	.578	.948
	t_{10}	Bias	−.009	.027	−.002	.147	−.333	.007
		Empirical variance	.033	.103	.022	.021	.039	.020
		Estimated variance	.033	.099	.019	.017	.037	.018
		95% cv	.950	.941	.934	.756	.586	.937
	t_5	Bias	−.075	.180	−.006	.120	−.272	.006
		Empirical variance	.038	.128	.023	.022	.042	.020
		Estimated variance	.039	.124	.020	.018	.040	.018
		95% cv	.952	.941	.935	.816	.729	.937
	Uniform	Bias	−.006	.021	−.003	.147	−.331	.007
		Empirical variance	.032	.099	.020	.021	.038	.019
		Estimated variance	.033	.098	.019	.017	.037	.018
		95% cv	.957	.945	.941	.768	.567	.951

NOTE: 95% CV represents the coverage probability of 95% confidence intervals.

expression, we can see that the error structure was much closer to normal based on the quantile–quantile plot in Figure 2. We also performed the Pearson chi-squared test, under which we obtained a p value $> .7$ for K ranging from 4 to 10; thus a normal error structure after the transformation was quite convincing. For patients with duplicated readings of biomarker expression, we used the averaged value $\bar{W}$ and the individual observations (W_1, W_2) for the analysis. Table 5 shows that ignoring the measurement error could cause severe bias, particularly in the estimates of the intercept and the biomarker expression effect. We found that biomarker expression significantly affected patient survival; a higher expression was associated with a shorter survival time. The naive method tended to underestimate the variance and produce a downward bias for the biomarker effect. As for other error-free covariates, the proposed method and the naive method yielded similar estimates. Patients with a tumor histology of adenocarcinoma had a significantly better survival rate than those with squamous cell carcinoma; moreover, younger patients could be expected to live longer at a lower risk of death. There was no significant difference in survival across sex in this study population, although there was a trend that women might live longer.

The cure threshold in our model could be determined through consultation with physicians, which is a medical issue based on the patient population and disease status. Because such a threshold is restricted to lie to the right of the largest failure time, we conducted a sensitivity analysis by taking the cure threshold at 7, 7.5, or 8 years. We found that the parameter estimates were not sensitive to the specification of the threshold, because it only affected the censored observations at the right tail.

6. DISCUSSION

We have proposed a semiparametric cure rate model with covariate measurement errors. The model inherits the well-known proportional hazards structure, with the corrected score functions derived based on NPMLE to estimate $\boldsymbol{\beta}$ and $F(t)$. The asymptotic consistency and root-n convergence of the estimators were established through modern empirical process techniques. Simulation studies showed that the corrected score approach produced consistent estimators, whereas the naive estimation typically led to severe biases.

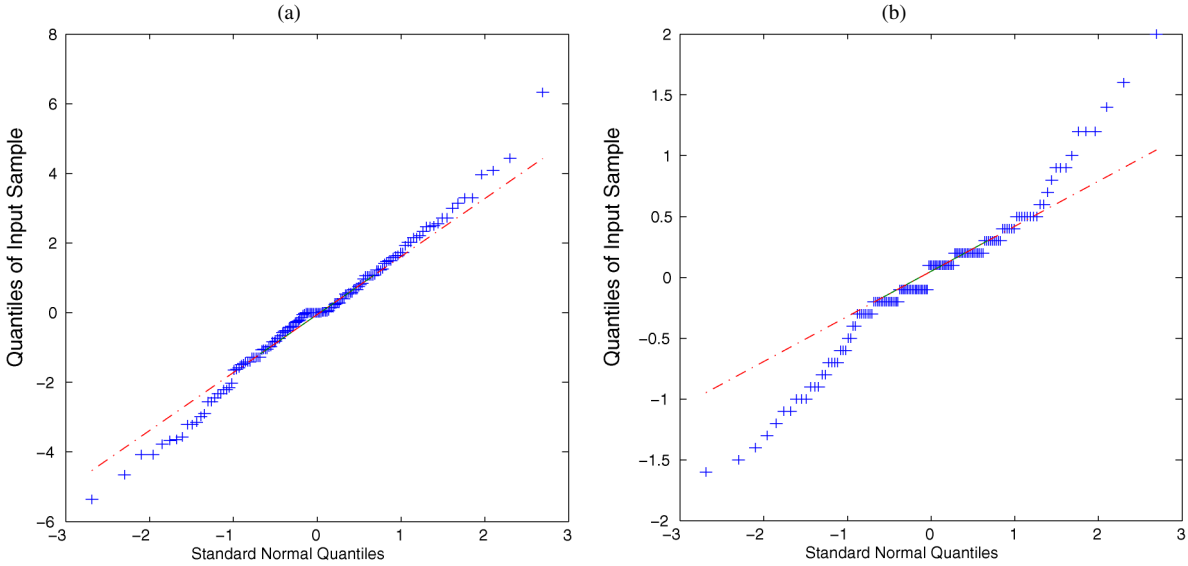

Figure 2. Quantile–quantile plots of the error with the transformed readings of biomarker expression (a) and the original observations (b) for the lung cancer data.

For the proportional hazard cure model, the NPMLE and partial likelihood estimator were in fact identical. This can be verified by assigning $n\lambda$ in NPMLE to $\sum_{Y_j=\infty} e^{\mathbf{X}_j^T\boldsymbol{\beta}}$. The details of this derivation are given in the Appendix. Naturally, the corrected score estimator thus also can be derived equivalently through the partial likelihood approach. In our view, using NPMLE has several advantages. First, consistency is self-evident through a conditional distribution argument. Second, the cure rate model and the Cox model are fundamentally different; therefore, to estimate the intercept b using the partial likelihood procedure, we need to reparameterize the model by taking $b = \log \Lambda_0(\infty)$. Third, the partial likelihood approach is restricted only to the proportional hazards structure, whereas NPMLE is much more general. For example, the proportional odds cure model can be easily handled by NPMLE, whereas the partial likelihood is not applicable (see Zeng et al. 2006). Even though finding a corrected score is not always straightforward or even possible, a general Monte Carlo–corrected score method can be implemented in practice. In this regard, $\mathbf{W}_i$ can be augmented with $\tilde{\mathbf{U}}_i\sqrt{-1}$ to form $\tilde{\mathbf{W}}_i$, where $\tilde{\mathbf{U}}_i$ has the same distribution as the measurement error $\mathbf{U}_i$; calculate the score function using $\tilde{\mathbf{W}}_i$; and set the real part to 0 to solve for $\boldsymbol{\beta}$. Although the Monte Carlo–corrected score approach does not always guarantee a consistent estimator, in the case when a true corrected score does exist, it will be consistent; thus it can be viewed as a numerical way of finding a corrected score. Further exploration in these areas should be worthwhile.

Table 5. Regression coefficient estimates and estimated variances for the lung cancer data, using the averaged and individual reading of biomarker expression

Estimate	Intercept	Histology	Age	Sex	Biomarker
Averaged reading $\bar{W}$					
Proposed estimate	−.2904	−.5622	.4352	−.0687	.0505
Estimated variance	.2990	.0639	.0138	.0450	.0077
Naive estimate	−.1672	−.5306	.4337	−.0748	.0299
Estimated variance	.1347	.0489	.0130	.0426	.0026
Individual reading (W_1, W_2)					
Proposed estimate	−.2648	−.5555	.4348	−.0697	.0462
Estimated variance	.2885	.0624	.0138	.0450	.0074
Naive estimate	−.1448	−.5248	.4334	−.0758	.0261
Estimated variance	.0923	.0476	.0134	.0434	.0020

The corrected-score method belongs to the family of functional approaches that make no distributional assumptions on the unobservable true covariates, as opposed to structural models that specify a distribution of $\mathbf{X}$. However, the corrected score and its generalized form depend on the additive error structure. Because in reality not all measurements can be transformed to normality and/or additivity, the more general semiparamatric approach proposed by Tsiatis and Ma (2004) is worth exploring. Preliminary studies toward this end have uncovered several modeling and computational issues, including the need to estimate the censoring mechanism or strong assumptions, such as censoring, independent of the covariates. These same difficulties also prevent us from absorbing the transformation inside of the estimation procedure itself.

An alternative approach in functional measurement error models is based on the simulation-extrapolation (SIMEX) method (Cook and Stefanski 1994; Stefanski and Cook 1995; Li and Lin 2003; Greene and Cai 2004). SIMEX first simulates data sets with an increasing amount of measurement errors and then extrapolates back to the nonerror case. Because the correct extrapolation function is generally unknown, in theory SIMEX can produce only approximately consistent estimates. However, in practice its performance is often satisfactory, and sometimes it even outperforms the asymptotically "correct" methods. It would be interesting to implement SIMEX under the cure rate models with measurement errors.

In contrast to the usual classical measurement error structure, where $\mathbf{W} = \mathbf{X} + \mathbf{U}$, $\mathbf{U}$ is independent of $\mathbf{X}$, in another class of

errors, called Berkson error, $\mathbf{X} = \mathbf{W} + \mathbf{U}$ and $\mathbf{U}$ is independent of $\mathbf{W}$. Berkson error models are often relatively easier to deal with, because the distribution of the unobservable variable $\mathbf{X}$ does not appear in the likelihood. In the Cox model framework, Zucker (2005) considered such an error structure and proposed a consistent estimator. Similar estimators can be developed for the cure rate model. Further consideration in the presence of a mixture of classical and Berkson errors may be also of interest.

APPENDIX: PROOFS

Proof of Theorem 1

Multiplying (6) by $\hat{p}_{(i)} = \hat{F}_n\{Y_i\}$ on both sides and summing over the m equations, we have

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} \Delta_i I(Y_i < \infty) - \int_0^{\infty} H_n(y, \hat{\boldsymbol{\beta}}_n)\, d\hat{F}_n(y), \tag{A.1}$$

where

$$H_n(y, \hat{\boldsymbol{\beta}}_n) = \frac{1}{n}\sum_{Y_j < \infty} I(Y_j \geq y) e^{\mathbf{W}_j^T \hat{\boldsymbol{\beta}}_n - \hat{\boldsymbol{\beta}}_n \mathbf{V} \hat{\boldsymbol{\beta}}_n/2}.$$

Thus $\hat{F}_n\{Y_i\} = \Delta_i / n\{\hat{\lambda}_n + H_n(Y_i, \hat{\boldsymbol{\beta}}_n)\}$. Obviously, from (A.1), $\hat{\lambda}_n$ should be bounded by a constant with probability 1; therefore, by choosing a subsequence, still indexed by $\{n\}$, we assume that $\hat{\lambda}_n \to \lambda^*$. By choosing a further subsequence, we assume that $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}^*$ and $\hat{F}_n \to F^*$ pointwise.

Note that the classes $\{\mathbf{W}\}$ and $\{I(Y \geq y)\}$ are P–Donsker, because of their monotonicity and uniform boundedness. Under the continuously differentiable operation of taking the exponential and the algebraic operation of multiplication, the class

$$\left\{ I(\infty > Y \geq y) e^{\mathbf{W}^T \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta}/2} : \boldsymbol{\beta} \in \mathcal{B}_0 \right\}$$

also is P–Donsker (see van der Vaart and Wellner 2000, thms. 2.7.5, 2.10.6, and 2.10.8) and thus is Glivenko–Cantelli. Due to the Glivenko–Cantelli theorem and the bounded convergence theorem, we conclude that uniformly in y, $H_n(y, \hat{\boldsymbol{\beta}}_n) \to H^*(y)$, where

$$H^*(y) = E\left\{ I(\infty > Y \geq y) e^{\mathbf{W}^T \boldsymbol{\beta}^* - \boldsymbol{\beta}^* \mathbf{V} \boldsymbol{\beta}^*/2} \right\}.$$

Moreover, the right side of (A.1) converges to

$$\lambda^* = E\{\Delta I(Y < \infty)\} - E\left\{ I(Y < \infty) \int_0^Y H^*(y)\, dF^*(y) \right\}.$$

We next show that $|\lambda^* + H^*(y)|$ is bounded away from 0. Because each $\hat{F}_n\{Y_i\}$ is nonnegative and $\sum_{i=1}^n \hat{F}_n\{Y_i\} = 1$, we have

$$1 = \sum_{i=1}^{n} \frac{I(Y_i < \infty)\Delta_i}{n(\hat{\lambda}_n + H_n(Y_i, \hat{\boldsymbol{\beta}}_n))} = \sum_{i=1}^{n} \frac{I(Y_i < \infty)\Delta_i}{n|\hat{\lambda}_n + H_n(Y_i, \hat{\boldsymbol{\beta}}_n)|}$$
$$\geq \frac{1}{n}\sum_{i=1}^{n} \frac{I(Y_i < \infty)\Delta_i}{|\hat{\lambda}_n + H_n(Y_i, \hat{\boldsymbol{\beta}}_n)| + \epsilon},$$

for any positive constant ϵ. Because $H_n(y, \hat{\boldsymbol{\beta}}_n)$ converges uniformly to $H^*(y)$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \frac{I(Y_i < \infty)\Delta_i}{|\hat{\lambda}_n + H_n(Y_i, \hat{\boldsymbol{\beta}}_n)| + \epsilon} - \frac{1}{n}\sum_{i=1}^{n} \frac{I(Y_i < \infty)\Delta_i}{|\lambda^* + H^*(Y_i)| + \epsilon} \to 0.$$

Then, after taking limits on both sides, we obtain $1 \geq E\{\Delta I(Y < \infty)/(|\lambda^* + H^*(Y)| + \epsilon)\}$. Let $\epsilon \to 0$; then we have

$$1 \geq \int_0^{\infty} \frac{c_0\, dy}{|\lambda^* + H^*(y)|}, \tag{A.2}$$

where c_0 is a positive constant. This implies that there exists a $\delta^* > 0$ such that $|\lambda^* + H^*(y)| > \delta^*$, because otherwise, we have that $\inf_y |\lambda^* + H^*(y)| = 0$. If $H^*(\infty) + \lambda^* = 0$, then $H^*(\infty) = -\lambda^* = 0$. In this case, $|\lambda^* + H^*(y)| = H^*(y) < 1$ for sufficiently large y, which contradicts (A.2). If $\lambda^* = -H^*(y_0)$ for a finite y_0, then (A.2) becomes $1 \geq c_0 \int_0^{\infty} 1/|H^*(y_0) - H^*(y)|\, dy$. This is impossible, because $H^*(y)$ is continuously differentiable in a neighborhood of y_0. Furthermore, $|\lambda^* + H^*(y)| > \delta^*$ implies that when n is large, $|\hat{\lambda}_n + H_n(y, \hat{\boldsymbol{\beta}}_n)| > \delta^*$. Note that $\hat{F}_n(y) = n^{-1}\sum_{i=1}^n \Delta_i I(Y_i \leq y)/|\hat{\lambda}_n + H_n(Y_i, \hat{\boldsymbol{\beta}}_n)|$ so $\hat{F}_n(y)$ converges uniformly to $F^*(y) = E\{\Delta I(Y \leq y)/|\lambda^* + H^*(Y)|\}$.

Based on the definitions of $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, F)$ and $S_F(\boldsymbol{\beta}, F)[\int Q_F(h_2)\, dF]$, we have

$$\mathbb{P}_n\left\{ \mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)^T \mathbf{h}_1 + S_F(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)\left[\int Q_{\hat{F}_n}(h_2)\, d\hat{F}_n \right] \right\} = 0$$

for any $(\mathbf{h}_1, h_2) \in \mathcal{H}$. Using the uniform convergence of $(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)$ to $(\boldsymbol{\beta}^*, F^*)$, as $n \to \infty$, we obtain

$$\mathbb{P}\left\{ \mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^*, F^*)^T \mathbf{h}_1 + S_F(\boldsymbol{\beta}^*, F^*)\left[\int Q_{F^*}(h_2)\, dF^* \right] \right\} = 0.$$

On the other hand, we also have

$$\mathbb{P}\left\{ \mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^T \mathbf{h}_1 + S_F(\boldsymbol{\beta}_0, F_0)\left[\int Q_{F_0}(h_2)\, dF_0 \right] \right\}$$
$$= \mathbb{P}\left\{ \mathbf{l}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^T \mathbf{h}_1 + l_F(\boldsymbol{\beta}_0, F_0)\left[\int Q_{F_0}(h_2)\, dF_0 \right] \right\}$$
$$= 0, \tag{A.3}$$

because of our construction. Based on the uniqueness assumption and continuity of F_0, we obtain $\boldsymbol{\beta}_0 = \boldsymbol{\beta}^*$ and $F^* = F_0$.

Proof of Theorem 2

Following the definition of $\mathcal{H}$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{F}_n - F_0)$ can be treated as a linear functional in the metric space $l^\infty(\mathcal{H})$, which is defined as

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{F}_n - F_0)(\mathbf{h}_1, h_2) = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T \mathbf{h}_1 + \sqrt{n}\int h_2(t)\, d(\hat{F}_n - F_0).$$

We next establish the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0, \hat{F}_n - F_0)$ in $l^\infty(\mathcal{H})$.

We denote the derivative of $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, F)$ with respect to $\boldsymbol{\beta}$ as $\mathbf{S}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}, F)$, the derivative of $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, F)$ with respect to F along the path $F_\epsilon = F + \epsilon(\hat{F}_n - F)$ as $\mathbf{S}_{\boldsymbol{\beta}F}(\boldsymbol{\beta}, F)[\hat{F}_n - F]$, the derivative of $S_F(\boldsymbol{\beta}, F)[\int Q_F(h_2)\, dF]$ with respect to $\boldsymbol{\beta}$ as $\mathbf{S}_{F\boldsymbol{\beta}}(\boldsymbol{\beta}, F) \times [\int Q_F(h_2)\, dF]$, and the derivative of $S_F(\boldsymbol{\beta}, F)[\int Q_F(h_2)\, dF]$ with respect to F along the path $F_\epsilon = F + \epsilon(\hat{F}_n - F)$ as $S_{FF}(\boldsymbol{\beta}, F) \times [\int Q_F(h_2)\, dF, \hat{F}_n - F]$. Explicitly, we have

$$\mathbf{S}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}, F) = F(Y) e^{\mathbf{W}^T \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta}/2} \{\mathbf{V} - (\mathbf{W} - \mathbf{V}\boldsymbol{\beta})(\mathbf{W} - \mathbf{V}\boldsymbol{\beta})^T\},$$

$$\mathbf{S}_{F\boldsymbol{\beta}}(\boldsymbol{\beta}, F)\left[\int Q_F(h_2)\, dF \right] = -\int_0^Y Q_F(h_2)\, dF(t)\, e^{\mathbf{W}^T \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta}/2} (\mathbf{W} - \mathbf{V}\boldsymbol{\beta}),$$

$$\mathbf{S}_{\boldsymbol{\beta}F}(\boldsymbol{\beta}, F)[\hat{F}_n - F] = -\{\hat{F}_n(Y) - F(Y)\} e^{\mathbf{W}^T \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta}/2} (\mathbf{W} - \mathbf{V}\boldsymbol{\beta}),$$

and

$$S_{FF}(\boldsymbol{\beta}, F)\Big[\int Q_F(h_2)\,dF, \hat{F}_n - F\Big]$$
$$= -e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}\int_0^Y Q_F(h_2)\,d(\hat{F}_n - F)$$
$$+ \{e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}F(Y) - I(Y<\infty)\Delta\}\int_0^\infty h_2(t)\,d(\hat{F}_n - F).$$

Our estimating equations require that

$$\mathbb{P}_n\Big\{\mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)^T\mathbf{h}_1 + S_F(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)\Big[\int Q_{\hat{F}_n}(h_2)\,d\hat{F}_n\Big]\Big\} = 0.$$

Noting (A.3), we obtain

$$\sqrt{n}(\mathbb{P}_n - \mathbb{P})\Big\{\mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)^T\mathbf{h}_1 + S_F(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)\Big[\int Q_{\hat{F}_n}(h_2)\,d\hat{F}_n\Big]\Big\}$$
$$= -\sqrt{n}\mathbb{P}\Big\{\mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)^T\mathbf{h}_1 + S_F(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)\Big[\int Q_{\hat{F}_n}(h_2)\,d\hat{F}_n\Big]\Big\}$$
$$+ \sqrt{n}\mathbb{P}\Big\{\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^T\mathbf{h}_1 + S_F(\boldsymbol{\beta}_0, F_0)\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big\}. \tag{A.4}$$

We now examine the left and right sides of (A.4). For the left side, we note that the classes $\{\mathbf{W}\}$ and $\{F(Y)\}$ are P–Donsker, because of their monotonicity and uniform boundedness. Under the continuously differentiable operation of taking the exponential, and the algebraic operation of multiplication, the class

$$\Big\{e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}, e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}(\mathbf{W}-\mathbf{V}\boldsymbol{\beta}),$$
$$e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})F(Y):$$
$$\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| < \delta_0, \sup_y|F(y)-F_0(y)| < \delta_0\Big\}$$

also is P–Donsker (see van der Vaart and Wellner 2000, thms. 2.7.5, 2.10.6, and 2.10.8). In addition, the class $\{h_2, Q_F(h_2), \int_0^Y Q_F(h_2)\,dF : \|h_2\|_V \le 1, \sup_y|F(y)-F_0(y)| < \delta_0\}$ contains functions of Y with bounded variations, and so also is P–Donsker. Therefore, from the explicit expressions of $\mathbf{S}_{\boldsymbol{\beta}}$ and S_F, the preservation of the Donsker classes under algebraic operations implies that the class

$$\mathcal{A} = \Big\{\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, F)^T\mathbf{h}_1 + S_F(\boldsymbol{\beta}, F)\Big[\int Q_F(h_2)\,dF\Big]:$$
$$\|\mathbf{h}_1\| \le 1, \|h_2\|_V \le 1, \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| + \sup_y|F(y)-F_0(y)| < \delta_0\Big\}$$

is P–Donsker. On the other hand, based on the consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{F}_n$, the bounded norm of $\mathbf{h}_1$ and the bounded total variation of h_2, it is straightforward to show that

$$\mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)^T\mathbf{h}_1 + S_F(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)\Big[\int Q_{\hat{F}_n}(h_2)\,d\hat{F}_n\Big]$$
$$\to \mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^T\mathbf{h}_1 + S_F(\boldsymbol{\beta}_0, F_0)\Big[\int Q_{F_0}(h_2)\,dF_0\Big]$$

uniformly in $(\mathbf{h}_1, h_2) \in \mathcal{H}$. Thus the left side of (A.4) is equal to

$$\sqrt{n}(\mathbb{P}_n - \mathbb{P})\Big\{\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^T\mathbf{h}_1 + S_F(\boldsymbol{\beta}_0, F_0)\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big\}$$
$$+ o_p(1),$$

where $o_p(1)$ is a random variable that converges to 0 in probability in $l^\infty(\mathcal{H})$. As a result, the left side of (A.4) converges weakly to a mean-0 Gaussian process in $l^\infty(\mathcal{H})$.

For the right side, simple algebra shows that uniformly in $(\mathbf{h}_1, h_2) \in \mathcal{H}$,

$$\Big|\mathbf{S}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)^T\mathbf{h}_1 + S_F(\hat{\boldsymbol{\beta}}_n, \hat{F}_n)\Big[\int Q_{\hat{F}_n}(h_2)\,d\hat{F}_n\Big]$$
$$- \mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)^T\mathbf{h}_1 - S_F(\boldsymbol{\beta}_0, F_0)\Big[\int Q_{F_0}(h_2)\,dF_0\Big]$$
$$- \Big\{(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\mathbf{S}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)\mathbf{h}_1$$
$$+ (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\mathbf{S}_{F\boldsymbol{\beta}}(\boldsymbol{\beta}_0, F_0)\Big[\int Q_{F_0}(h_2)\,dF_0\Big]$$
$$+ \mathbf{h}_1^T\mathbf{S}_{\boldsymbol{\beta}F}[\hat{F}_n - F_0]$$
$$+ S_{FF}(\boldsymbol{\beta}_0, F_0)\Big[\int Q_{F_0}(h_2)\,dF_0, \hat{F}_n - F_0\Big]\Big\}\Big|$$
$$\le o_p\{\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \|\hat{F}_n - F_0\|_{l^\infty(\mathcal{H})}\}.$$

Thus, combining this with the expressions of $\mathbf{S}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \mathbf{S}_{\boldsymbol{\beta}F}, \mathbf{S}_{F\boldsymbol{\beta}}$, and S_{FF}, we obtain that the right side of (A.4) equals

$$-\sqrt{n}\Big\{(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^T\boldsymbol{\Omega}_{\boldsymbol{\beta}}(\mathbf{h}_1, Q_{F_0}(h_2))$$
$$+ \int_0^\infty \Omega_F(\mathbf{h}_1, Q_{F_0}(h_2))\,d(\hat{F}_n - F_0)(y)\Big\}$$
$$+ o_p\{\sqrt{n}(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| + \|\hat{F}_n - F_0\|_{l^\infty(\mathcal{H})})\},$$

where

$$\boldsymbol{\Omega}_{\boldsymbol{\beta}}(\mathbf{h}_1, Q_{F_0}(h_2))$$
$$= E[F_0(Y)e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}\{\mathbf{V} - (\mathbf{W}-\mathbf{V}\boldsymbol{\beta})(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})^T\}\mathbf{h}_1]$$
$$- E\Big[e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})\int_0^Y Q_{F_0}(h_2)\,dF_0(t)\Big]$$

and

$$\Omega_F(\mathbf{h}_1, Q_{F_0}(h_2))$$
$$= E\{e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}F_0(Y) - I(Y<\infty)\Delta\}Q_{F_0}\{h_2(y)\}$$
$$- E\big(I(y\le Y)e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}[(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})^T\mathbf{h}_1 + Q_{F_0}\{h_2(y)\}]\big)$$
$$+ \int_0^\infty E\big(I(y\le Y)e^{\mathbf{W}^T\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}/2}$$
$$\times[(\mathbf{W}-\mathbf{V}\boldsymbol{\beta})^T\mathbf{h}_1 + Q_{F_0}\{h_2(y)\}]\big)\,dF_0(y).$$

Note that the last term in the foregoing expression is added to ensure that $\int \Omega_F(\mathbf{h}_1, Q_{F_0}(h_2))\,dF_0 = 0$. We now show that $(\boldsymbol{\Omega}_{\boldsymbol{\beta}}, \Omega_F)$, which is a linear operator on a subspace of $\mathcal{H}$ to itself, with mean-0 Ω_F under F_0, is invertible. Note that

$$\boldsymbol{\Omega}_{\boldsymbol{\beta}}(\mathbf{h}_1, Q_{F_0}(h_2))$$
$$= E\Big\{\mathbf{S}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{h}_1 + \mathbf{S}_{F\boldsymbol{\beta}}\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big\}$$
$$= E\Big[E\Big\{\frac{\partial\mathbf{S}_{\boldsymbol{\beta}}}{\partial\boldsymbol{\beta}}\mathbf{h}_1 + \frac{\partial S_F[\int Q_{F_0}(h_2)\,dF_0]}{\partial\boldsymbol{\beta}}\Big|\mathbf{X}, Y, \Delta\Big\}\Big]$$
$$= E\Big\{\frac{\partial E(\mathbf{S}_{\boldsymbol{\beta}}|\mathbf{X}, Y, \Delta)}{\partial\boldsymbol{\beta}}\mathbf{h}_1 + \frac{\partial E(S_F[\int Q_{F_0}(h_2)\,dF_0]|\mathbf{X}, Y, \Delta)}{\partial\boldsymbol{\beta}}\Big\}$$
$$= E\Big(\mathbf{l}_{\boldsymbol{\beta}\boldsymbol{\beta}}\mathbf{h}_1 + \mathbf{l}_{F\boldsymbol{\beta}}\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big)$$

$$= -E\Big(\mathbf{l}_{\boldsymbol\beta}\mathbf{l}_{\boldsymbol\beta}^T\mathbf{h}_1 + l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\mathbf{l}_{\boldsymbol\beta}\Big)$$

and, similarly,

$$\begin{aligned}
&\int_0^\infty \Omega_F\big(\mathbf{h}_1, Q_{F_0}(h_2)\big)\,dG(y)\\
&\quad= E\Big(\mathbf{h}_1^T\mathbf{S}_{\boldsymbol\beta F}[G] + S_{FF}\Big[\int Q_{F_0}(h_2)\,dF_0, G\Big]\Big)\\
&\quad= E\Big\{E\Big(\mathbf{h}_1^T\mathbf{S}_{\boldsymbol\beta F}[G] + S_{FF}\Big[\int Q_{F_0}(h_2)\,dF_0, G\Big]\Big|\mathbf{X}, Y, \Delta\Big)\Big\}\\
&\quad= E\Big(\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta F}[G] + l_{FF}\Big[\int Q_{F_0}(h_2)\,dF_0, G\Big]\Big)\\
&\quad= -E\Big(\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta} l_F[G] + l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big] l_F[G]\Big)
\end{aligned}$$

for an arbitrary function G such that $\int_0^\infty dG(y) = 0$. In particular, for $G = \int Q_{F_0}(h_2)\,dF_0$, we obtain

$$\begin{aligned}
&\int_0^\infty \Omega_F\big(\mathbf{h}_1, Q_{F_0}(h_2)\big) Q_{F_0}(h_2)\,dF_0\\
&\quad= -E\Big\{\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta} l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big] + \Big(l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big)^2\Big\}.
\end{aligned}$$

If for certain $(\mathbf{h}_1, h_2)$ and an arbitrary G satisfying $\int_0^\infty dG(y) = 0$, we have

$$\boldsymbol\Omega_{\boldsymbol\beta}\big(\mathbf{h}_1, Q_{F_0}(h_2)\big) = \mathbf{0} \qquad \text{and}$$

$$\int_0^\infty \Omega_F\big(\mathbf{h}_1, Q_{F_0}(h_2)\big)\,dG(y) = 0,$$

then we obtain

$$\begin{aligned}
0 &= \mathbf{h}_1^T\boldsymbol\Omega_{\boldsymbol\beta}\big(\mathbf{h}_1, Q_{F_0}(h_2)\big) + \int_0^\infty \Omega_F\big(\mathbf{h}_1, Q_{F_0}(h_2)\big) Q_{F_0}(h_2)\,dF_0\\
&= -\mathbf{h}_1^T E\Big(\mathbf{l}_{\boldsymbol\beta}\mathbf{l}_{\boldsymbol\beta}^T\mathbf{h}_1 + l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\mathbf{l}_{\boldsymbol\beta}\Big)\\
&\quad - E\Big\{\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta} l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big] + \Big(l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big)^2\Big\}\\
&= -E\Big\{(\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta})^2 + 2\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta} l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\\
&\qquad + \Big(l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big)^2\Big\}\\
&= -E\Big\{\Big(\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta} + l_F\Big[\int Q_{F_0}(h_2)\,dF_0\Big]\Big)^2\Big\};
\end{aligned}$$

therefore, with probability 1, $\mathbf{h}_1^T\mathbf{l}_{\boldsymbol\beta} + l_F[\int Q_{F_0}(h_2)\,dF_0] = 0$. When $Y = \infty$, $l_F[\int Q_{F_0}(h_2)\,dF_0] = 0$, and thus we obtain $\mathbf{h}_1 = \mathbf{0}$. Subtracting l_F evaluated at $(Y < \infty, \Delta = 1)$ by its value at $(Y < \infty, \Delta = 0)$, we obtain $l_F[\int Q_{F_0}(h_2)\,dF_0](Y, \Delta = 1) - l_F[\int Q_{F_0}(h_2)\,dF_0](Y, \Delta = 0) = Q_{F_0}\{h_2(Y)\} = 0$ for any $Y < \infty$, and thus $h_2 = 0$. Therefore, we have shown that $(\boldsymbol\Omega_{\boldsymbol\beta}, \Omega_F)$ is indeed invertible. Denoting its inverse $(\boldsymbol\Omega_{\boldsymbol\beta}^{-1}, \Omega_F^{-1})$, we can rewrite (A.4) as

$$\begin{aligned}
&\sqrt{n}\Big\{(\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0)^T\mathbf{h}_1 + \int_0^\infty Q_{F_0}(h_2)\,d(\hat F_n - F_0)\Big\}\\
&\quad= -\sqrt{n}(\mathbb{P}_n - \mathbb{P})\Big\{\mathbf{S}_{\boldsymbol\beta}(\boldsymbol\beta_0, F_0)^T\boldsymbol\Omega_{\boldsymbol\beta}^{-1}\big(\mathbf{h}_1, Q_{F_0}(h_2)\big)\\
&\qquad + S_F(\boldsymbol\beta_0, F_0)\Big[\int \Omega_F^{-1}\big(\mathbf{h}_1, Q_{F_0}(h_2)\big)\,dF_0\Big]\Big\}\\
&\qquad + o_p\big\{\sqrt{n}(\|\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0\| + \|\hat F_n - F_0\|_{l^\infty})\big\} + o_p(1)
\end{aligned}$$

for all $(\mathbf{h}_1, h_2) \in \mathcal{H}$. So, $\sqrt{n}\{(\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0)^T\mathbf{h}_1 + \int_0^\infty Q_{F_0}(h_2)\,d(\hat F_n - F_0)\} = O_p(1)$ for all $(\mathbf{h}_1, h_2) \in \mathcal{H}$, and thus $\{\sqrt{n}(\|\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0\| + \|\hat F_n - F_0\|_{l^\infty})\} = O_p(1)$. As an immediate result,

$$\begin{aligned}
&\sqrt{n}\Big\{(\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0)^T\mathbf{h}_1 + \int_0^\infty Q_{F_0}(h_2)\,d(\hat F_n - F_0)\Big\}\\
&\quad= -\sqrt{n}(\mathbb{P}_n - \mathbb{P})\Big\{\mathbf{S}_{\boldsymbol\beta}(\boldsymbol\beta_0, F_0)^T\boldsymbol\Omega_{\boldsymbol\beta}^{-1}\big(\mathbf{h}_1, Q_{F_0}(h_2)\big)\\
&\qquad + S_F(\boldsymbol\beta_0, F_0)\Big[\int \Omega_F^{-1}\big(\mathbf{h}_1, Q_{F_0}(h_2)\big)\,dF_0\Big]\Big\}\\
&\qquad + o_p(1),
\end{aligned}$$

and thus the proof of the theorem is complete.

Proof of Theorem 3

From the proof of Theorem 2, if we take $Q_{F_0}(h_2) = 0$, then we obtain that

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0)^T\mathbf{h}_1\\
&\quad= -\sqrt{n}(\mathbb{P}_n - \mathbb{P})\Big\{\mathbf{S}_{\boldsymbol\beta}(\boldsymbol\beta_0, F_0)^T\boldsymbol\Omega_{\boldsymbol\beta}^{-1}(\mathbf{h}_1, 0)\\
&\qquad + S_F(\boldsymbol\beta_0, F_0)\Big[\int \Omega_F^{-1}(\mathbf{h}_1, 0)\,dF_0\Big]\Big\} + o_p(1).
\end{aligned}$$

Write

$$\begin{aligned}
\mathbf{b}_1(y) &= E\big\{I(y \le Y)e^{\mathbf{W}^T\boldsymbol\beta - \boldsymbol\beta^T\mathbf{V}\boldsymbol\beta/2}(\mathbf{W} - \mathbf{V}\boldsymbol\beta)\big\},\\
\mathbf{b}_2 &= \int_0^\infty \mathbf{b}_1(y)\,dF_0(y),\\
c_1 &= E\big\{e^{\mathbf{W}^T\boldsymbol\beta - \boldsymbol\beta^T\mathbf{V}\boldsymbol\beta/2}F_0(Y) - I(Y < \infty)\Delta\big\},\\
c_2(y) &= E\big\{I(y \le Y)e^{\mathbf{W}^T\boldsymbol\beta - \boldsymbol\beta^T\mathbf{V}\boldsymbol\beta/2}\big\},\\
\mathbf{b}_3 &= \Big[\int_0^\infty \{c_1 - c_2(y)\}^{-1}\,dF_0(y)\Big]^{-1}\\
&\qquad \times \Big[\int_0^\infty \{\mathbf{b}_1(y) - \mathbf{b}_2\}/\{c_1 - c_2(y)\}\,dF_0(y)\Big],\\
\mathbf{b}_4(y) &= \{c_1 - c_2(y)\}^{-1}\{\mathbf{b}_1(y) - \mathbf{b}_2 - \mathbf{b}_3\},\\
\mathbf{a}_1 &= E\big[F_0(Y)e^{\mathbf{W}^T\boldsymbol\beta - \boldsymbol\beta^T\mathbf{V}\boldsymbol\beta/2}\{\mathbf{V} - (\mathbf{W} - \mathbf{V}\boldsymbol\beta)(\mathbf{W} - \mathbf{V}\boldsymbol\beta)^T\}\big],\\
\mathbf{a}_2 &= E\Big[e^{\mathbf{W}^T\boldsymbol\beta - \boldsymbol\beta^T\mathbf{V}\boldsymbol\beta/2}(\mathbf{W} - \mathbf{V}\boldsymbol\beta)\int_0^Y \mathbf{b}_4(y)^T\,dF_0(y)\Big].
\end{aligned} \tag{A.5}$$

It can be verified that

$$\boldsymbol\Omega_{\boldsymbol\beta}^{-1}(\mathbf{h}_1, 0) = (\mathbf{a}_1 - \mathbf{a}_2)^{-1}\mathbf{h}_1 = (\mathbf{A}^{-1})^T\mathbf{h}_1$$

and

$$\Omega_F^{-1}(\mathbf{h}_1, 0) = \mathbf{b}_4(y)^T(\mathbf{A}^{-1})^T\mathbf{h}_1,$$

where $\mathbf{A} = (\mathbf{a}_1 - \mathbf{a}_2)^T$. Thus

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta_0)^T\mathbf{h}_1\\
&\quad= -\sqrt{n}(\mathbb{P}_n - \mathbb{P})\Big[\Big\{\mathbf{S}_{\boldsymbol\beta}(\boldsymbol\beta_0, F_0)^T + S_F(\boldsymbol\beta_0, F_0)\Big[\int \mathbf{b}_4(y)^T\,dF_0\Big]\Big\}\\
&\qquad \times (\mathbf{A}^{-1})^T\mathbf{h}_1\Big] + o_p(1).
\end{aligned}$$

Denote

$$\mathbf{B} = \Big\{\mathbf{S}_{\boldsymbol\beta}(\boldsymbol\beta_0, F_0) + S_F(\boldsymbol\beta_0, F_0)\Big[\int \mathbf{b}_4(y)\,dF_0\Big]\Big\}^{\otimes 2},$$

where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for a vector $\mathbf{a}$. Then the result follows.

Equivalence of the NPMLE and Partial Likelihood Estimator

The partial likelihood estimator for the proportional hazards cure rate model can be obtained by solving

$$\sum_{i=1}^{n} \Delta_i I(Y_i < \infty)\left\{\tilde{\mathbf{X}}_i - \frac{\sum_{j\in R_i} \tilde{\mathbf{X}}_j e^{\tilde{\mathbf{X}}_j^T\tilde{\boldsymbol{\beta}}}}{\sum_{j\in R_i} e^{\tilde{\mathbf{X}}_j^T\tilde{\boldsymbol{\beta}}}}\right\} = \mathbf{0} \qquad \text{(A.6)}$$

and

$$\sum_{i=1}^{n} \frac{\Delta_i I(Y_i < \infty)}{\sum_{j\in R_i} e^{\tilde{\mathbf{X}}_j^T\tilde{\boldsymbol{\beta}}}} - e^b = 0, \qquad \text{(A.7)}$$

where R_i is the risk set at Y_i and $\tilde{\mathbf{X}}_i$ and $\tilde{\boldsymbol{\beta}}$ represent $\mathbf{X}$ and $\boldsymbol{\beta}$ with the first component (corresponding to the intercept b) excluded (see, e.g., Tsodikov 1998b).

In the NPMLE, setting the Lagrange multiplier $n\lambda = \sum_{Y_j=\infty} e^{\mathbf{X}_j^T\boldsymbol{\beta}}$, from (3), we immediately obtain

$$p_i = \frac{\Delta_i I(Y_i < \infty)}{\sum_{j\in R_i} e^{\mathbf{X}_j^T\boldsymbol{\beta}}} \qquad \text{and} \qquad F(Y_i) = \sum_{Y_k \le Y_i} \frac{\Delta_k I(Y_k < \infty)}{\sum_{j\in R_k} e^{\mathbf{X}_j^T\boldsymbol{\beta}}},$$

and thus

$$\sum_{i=1}^{n} e^{\mathbf{X}_i^T\boldsymbol{\beta}}\mathbf{X}_i F(Y_i) = \sum_{i=1}^{n}\sum_{Y_k\le Y_i} \frac{e^{\mathbf{X}_i^T\boldsymbol{\beta}}\mathbf{X}_i \Delta_k I(Y_k < \infty)}{\sum_{j\in R_k} e^{\mathbf{X}_j^T\boldsymbol{\beta}}}$$

$$= \sum_{k=1}^{n} \Delta_k I(Y_k < \infty) \frac{\sum_{j\in R_k} e^{\mathbf{X}_j^T\boldsymbol{\beta}}\mathbf{X}_j}{\sum_{j\in R_k} e^{\mathbf{X}_j^T\boldsymbol{\beta}}},$$

and, from (5), we obtain

$$\sum_{i=1}^{n} \Delta_i I(Y_i < \infty)\left(\mathbf{X}_i - \frac{\sum_{j\in R_i} e^{\mathbf{X}_j^T\boldsymbol{\beta}}\mathbf{X}_j}{\sum_{j\in R_i} e^{\mathbf{X}_j^T\boldsymbol{\beta}}}\right),$$

equivalent to (A.6). From (4), we have

$$\sum_{i=1}^{n} \frac{\Delta_i I(Y_i < \infty)}{\sum_{j\in R_i} e^{\mathbf{X}_j^T\boldsymbol{\beta}}} = 1,$$

equivalent to (A.7).

[Received September 2007. Revised February 2008.]

REFERENCES

Augustin, T. (2004), "An Exact Corrected Log-Likelihood Function for Cox's Proportional Hazards Model Under Measurement Error and Some Extensions," *Scandinavian Journal of Statistics*, 31, 43–50.

Berkson, J., and Gage, R. P. (1952), "Survival Curve for Cancer Patients Following Treatment," *Journal of the American Statistical Association*, 47, 501–515.

Buzas, J. S. (1998), "Unbiased Scores in Proportional Hazards Regression With Covariate Measurement Error," *Journal of Statistical Planning & Inference*, 67, 247–257.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), London: CRC Press.

Chen, M. H., Ibrahim, J. G., and Sinha, D. (1999), "A New Bayesian Model for Survival Data With a Surviving Fraction," *Journal of the American Statistical Association*, 94, 909–919.

Cheng, S. C., and Wang, N. (2001), "Linear Transformation Models for Failure Time Data With Covariate Measurement Error," *Journal of the American Statistical Association*, 96, 706–716.

Cook, J. R., and Stefanski, L. A. (1994), "Simulation-Extrapolation Estimation in Parametric Measurement Error Models," *Journal of the American Statistical Association*, 89, 1314–1328.

Cox, D. R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

Eckert, R. S., Carroll, R. J., and Wang, N. (1997), "Transformations to Additivity in Measurement Error Models," *Biometrics*, 53, 262–272.

Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley.

Gorfine, M., Hsu, L., and Prentice, R. L. (2004), "Nonparametric Correction for Covariate Measurement Error in a Stratified Cox Model," *Biostatistics*, 5, 75–87.

Gray, R. J., and Tsiatis, A. A. (1989), "A Linear Rank Test for Use When the Main Interest Is in Differences in Cure Rates," *Biometrics*, 45, 899–904.

Greene, W. F., and Cai, J. (2004), "Measurement Error in Covariates in the Marginal Hazards Model for Multivariate Failure Time Data," *Biometrics*, 60, 987–996.

Hu, C., and Lin, D. Y. (2002), "Cox Regression With Covariate Measurement Error," *Scandinavian Journal of Statistics*, 29, 637–655.

——— (2004), "Semiparametric Failure Time Regression With Replicates of Mismeasured Covariates," *Journal of the American Statistical Association*, 99, 105–118.

Hu, P., Tsiatis, A. A., and Davidian, M. (1998), "Estimating the Parameters in the Cox Model When Covariate Variables Are Measured With Error," *Biometrics*, 54, 1407–1419.

Huang, Y., and Wang, C. Y. (2000), "Cox Regression With Accurate Covariates Unascertainable: A Nonparametric-Correction Approach," *Journal of the American Statistical Association*, 95, 1209–1219.

Kong, F. H., and Gu, M. (1999), "Consistent Estimation in the Cox Proportional Hazards Model With Covariate Measurement Errors," *Statistica Sinica*, 9, 953–970.

Kuk, A. Y. C., and Chen, C. H. (1992), "A Mixture Model Combining Logistic Regression With Proportional Hazards Regression," *Biometrika*, 79, 531–541.

Kulich, M., and Lin, D. Y. (2000), "Additive Hazards Regression With Covariate Measurement Error," *Journal of the American Statistical Association*, 95, 238–248.

Laska, E. M., and Meisner, M. J. (1992), "Nonparametric Estimation and Testing in a Cure Rate Model," *Biometrics*, 48, 1223–1234.

Li, Y., and Lin, X. (2000), "Covariate Measurement Errors in Frailty Models for Clustered Survival Data," *Biometrika*, 87, 849–866.

——— (2003), "Functional Inference in Frailty Measurement Error Models for Clustered Survival Data Using the SIMEX Approach," *Journal of the American Statistical Association*, 98, 191–203.

Li, Y., and Ryan, L. (2004), "Survival Analysis With Heterogeneous Covariate Measurement Error," *Journal of the American Statistical Association*, 99, 724–735.

Li, Y., Tiwari, R., and Guha, S. (2007), "Mixture Cure Survival Models With Dependent Censoring," *Journal of the Royal Statistical Society*, Ser. B, 69, 285–306.

Lu, W., and Ying, Z. (2004), "On Semiparametric Transformation Cure Models," *Biometrika*, 91, 331–343.

Maller, R., and Zhou, X. (1996), *Survival Analysis With Long-Term Survivors*, New York: Wiley.

Nakamura, T. (1990), "Corrected Score Function for Errors-in-Variables Models: Methodology and Application to Generalized Linear Models," *Biometrika*, 77, 127–137.

——— (1992), "Proportional Hazards Model With Covariates Subject to Measurement Error," *Biometrics*, 48, 829–838.

Nusser, S. M., Carriquiry, A. L., Dodd, K. W., and Fuller, W. A. (1996), "A Semiparametric Transformation Approach to Estimating Usual Daily Intake Distributions," *Journal of the American Statistical Association*, 91, 1440–1449.

Prentice, R. L. (1982), "Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model," *Biometrika*, 69, 331–342.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: Wiley.

Song, X., and Huang, Y. (2005), "On Corrected Score Approach for Proportional Hazards Model With Covariate Measurement Error," *Biometrics*, 61, 702–714.

Song, X., Davidian, M., and Tsiatis, A. A. (2002), "An Estimator for the Proportional Hazards Model With Multiple Longitudinal Covariates Measured With Error," *Biostatistics*, 3, 511–528.

Sposto, R., Sather, H. N., and Baker, S. A. (1992), "A Comparison of Tests of the Difference in the Proportion of Patients Who Are Cured," *Biometrics*, 48, 87–99.

Stefanski, L. A. (1989), "Unbiased Estimation of a Nonlinear Function of a Normal Mean With Application to Measurement Error Models," *Communications in Statistics, Part A—Theory and Methods,* 18, 4335–4358.

Stefanski, L. A., and Cook, J. R. (1995), "Simulation-Extrapolation: The Measurement Error Jackknife," *Journal of the American Statistical Association*, 90, 1247–1256.

Sy, J. P., and Taylor, J. M. G. (2000), "Estimation in a Cox Proportional Hazards Cure Model," *Biometrics*, 56, 227–236.

Tsiatis, A. A., and Davidian, M. (2001), "A Semiparametric Estimator for the Proportional Hazards Model With Longitudinal Covariates Measured With Error," *Biometrika*, 88, 447–458.

Tsiatis, A. A., and Ma, Y. (2004), "Locally Efficient Semiparametric Estimators for Functional Measurement Error Models," *Biometrika*, 91, 835–848.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995), "Modeling the Relationship of Survival to Longitudinal Data Measured With Error: Applications to Survival and CD4 Counts in Patients With AIDS," *Journal of the American Statistical Association*, 90, 27–37.

Tsodikov, A. (1998a), "A Proportional Hazards Model Taking Account of Long-Term Survivors," *Biometrics*, 54, 1508–1516.

——— (1998b), "Asymptotic Efficiency of a Proportional Hazards Model With Cure," *Statistics and Probability Letters*, 39, 237–244.

Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003), "Estimating Cure Rates From Survival Data: An Alternative to Two-Component Mixture Models," *Journal of the American Statistical Association*, 98, 1063–1078.

van der Vaart, A. W., and Wellner, J. A. (2000), *Weak Convergence and Empirical Processes*, New York: Springer.

Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997), "Regression Calibration in Failure Time Regression," *Biometrics*, 53, 131–145.

Wulfsohn, M. S., and Tsiatis, A. A. (1997), "A Joint Model for Survival and Longitudinal Data Measured With Error," *Biometrics*, 53, 330–339.

Yakovlev, A. Y., and Tsodikov, A. D. (1996), *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, Hackensack, NJ: World Scientific.

Zeng, D., Yin, G., and Ibrahim, J. G. (2006), "Semiparametric Transformation Models for Survival Data With a Cure Fraction," *Journal of the American Statistical Association*, 101, 670–684.

Zhou, H., and Pepe, M. S. (1995), "Auxiliary Covariate Data in Failure Time Regression," *Biometrika*, 82, 139–149.

Zhou, H., and Wang, C.-Y. (2000), "Failure Time Regression With Continuous Covariates Measured With Error," *Journal of the Royal Statistical Society*, Ser. B, 62, 657–665.

Zucker, D. M. (2005), "A Pseudo-Partial Likelihood Method for Semiparametric Survival Regression With Covariate Errors," *Journal of the American Statistical Association*, 100, 1264–1277.